Statistics: Lecture 2 - Random Variables and (Discrete) Probability Distributions

Jan Bauer

jan.bauer@dhbw-mannheim.de

20.11.19

(Soong 3-3.2, 4, 4.1, 6., 6.1)

DHBW Duale Hochschule Baden-Württemberg Mannheim

Table of Contents

DHBW
Duale Hochschule
Baden-Württemberg
Mannheim

Table of Contents

Discrete vs. Continuous Random Variables

- Discrete RV: Takes on only a discrete set of values
  - Example: The values 1, 2, 3,...
- Continuous RV: Takes on a continuum of possible values
  - Example: All values in the interval $[0, 1]$

Cumulated Distribution Function

Consider a rv X.

- The cumulative distribution function (CDF) provides the probability that X will be at or below any given value x. We define it by

$$F_X(x) \equiv \mathrm{P}(X \leq x) \; .$$

CDF Properties

- The CDF exists both for discrete and for continuous rvs
- $F_X(x) \in [0, 1]$ (*necessary* condition)
- $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$ (*necessary* condition)
- The CDF is non-decreasing (*necessary* condition)

Discrete Random Variable CDF Example

Let a discrete rv X assume values $-1, 1, 2$ and 3 with probabilities $0.25, 0.15, 0.2$ and 0.4, respectively. We then have

$$F_X(x) = Prob(X \leq x) = \begin{cases} 0 & x < -1 \\ 0.25 & -1 \leq x < 1 \\ 0.4 & 1 \leq x < 2 \\ 0.6 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

- $F_X(x) \in [0, 1]$ ✓
- $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$ ✓
- The CDF is non-decreasing ✓

Discrete Random Variable CDF Example

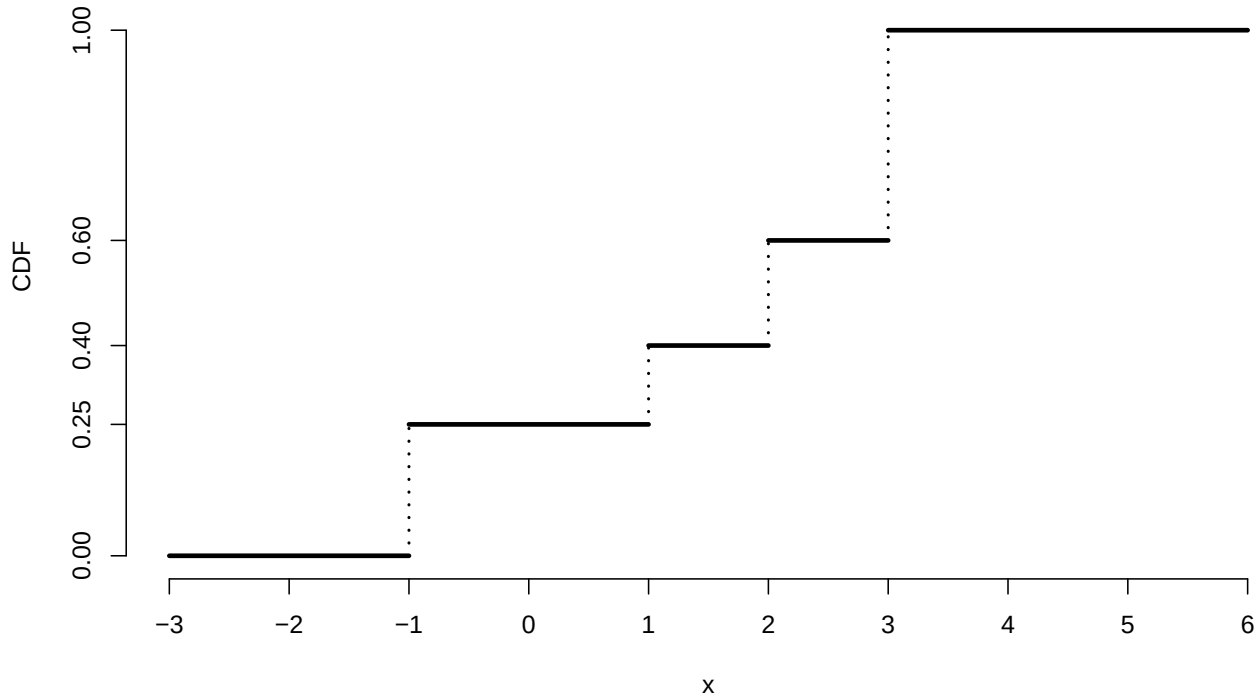

Figure: CDF for X

Discrete Random Variables pmf

For now on, we consider a discrete random variable X associated with the distinct outcomes x_i, $i = 1, 2, \ldots$.

- The function

$$f_X(x) \equiv \mathrm{P}(X = x) \quad \forall x$$

  is defined as the probability mass function (pmf) of X.

- What we might have expected already:

$$0 < f_X(x_i) \leq 1 \; \forall i \quad \text{and} \quad \sum_i f_X(x_i) = 1 \; .$$

Note: This are *necessary* conditions for a pmf

Discrete Random Variables pmf Example (1)

Let a discrete rv X assume values $-1, 1, 2$ and 3 with probabilities $0.25, 0.15, 0.2$ and 0.4, respectively. Is the pmf a valid one and what is the CDF?

$$f_X(x) = \begin{cases} f_X(-1) = 0.25 & x = -1 \\ f_X(1) = 0.15 & x = 1 \\ f_X(2) = 0.2 & x = 2 \\ f_X(3) = 0.4 & x = 3 \end{cases}.$$

Consider $x_1 = -1, x_2 = 1, x_3 = 2$ and $x_4 = 3$. Then

- $0 < f_X(x_i) \leq 1 \ \forall i$ ✓
- $\sum_i f_X(x_i) = f_X(-1) + f_X(1) + f_X(2) + f_X(3)$
  $= 0.25 + 0.15 + 0.2 + 0.4 = 1$ ✓
- CDF: ✓

Discrete Random Variable pmf Example (1)

Let a discrete rv X assume values $-1, 1, 2$ and 3 with probabilities $0.25, 0.15, 0.2$ and 0.4, respectively. We then have

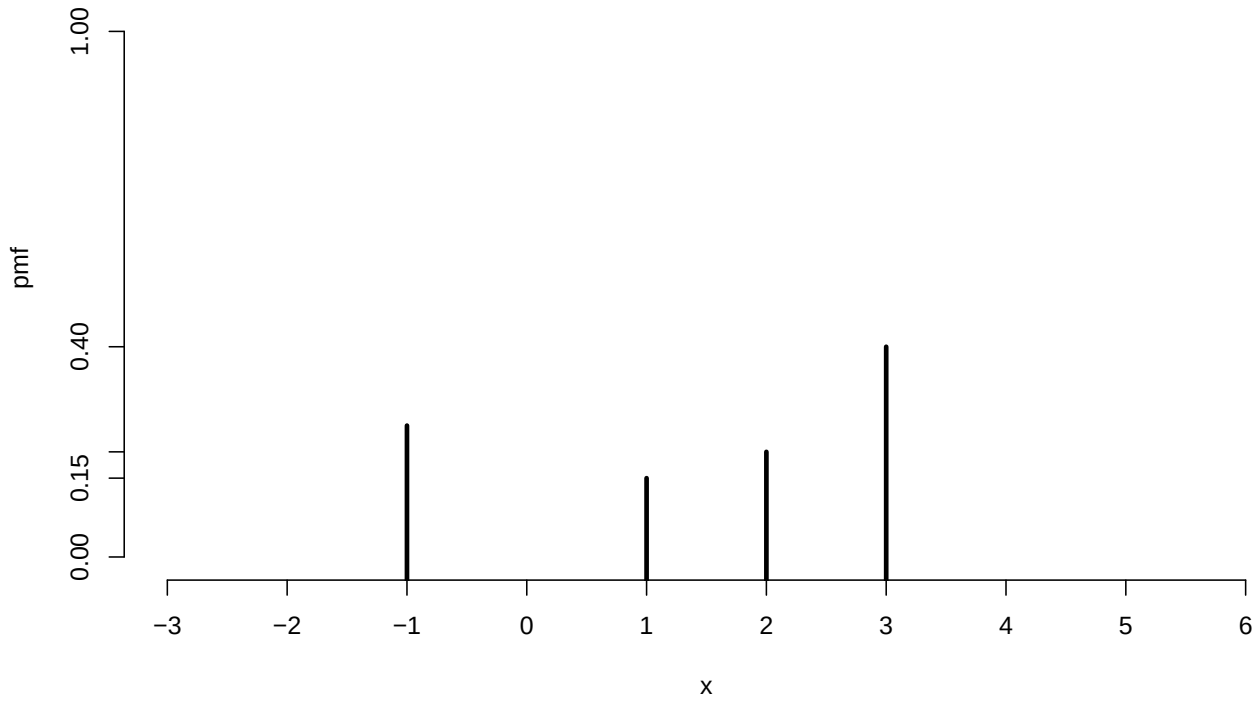

Figure: PMF for X

Discrete Random Variables pmf Example (2)

Consider you have a nightmare where you find a weird-looking coin-shaped item. Since you are a math enthusiast, you realise immediately that one side, $x = 1$, occurs with probability 0.4 and the other side, $x = 2$, with probability 0.45, respectively. Is the pmf a valid one and what is the CDF?

$$f_X(x) = \begin{cases} 0.4 & x = 1 \\ 0.45 & x = 2 \end{cases}.$$

It holds that

- $0 < f_X(x) \leq 1$ for $x = 1, 2$ ✓
- $\sum f_X(x) = f_X(1) + f_X(2) = 0.4 + 0.45 = 0.85 \neq 1$ ✗

Relations between discrete CDF and pmf

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \ .$$

$$F_X(x) = \sum_{i}^{i:x_i \leq x} f_X(x_i) \ .$$

$\rightsquigarrow$ CDF and pmf contain the same information (for discrete rvs)

Discrete Random Variables pmf Example (3)

Consider again the pmf given in Example (1):

$$f_X(x) = \begin{cases} f_X(-1) = 0.25 & x = -1 \\ f_X(1) = 0.15 & x = 1 \\ f_X(2) = 0.2 & x = 2 \\ f_X(3) = 0.4 & x = 3 \end{cases}.$$

What is $\mathrm{P}(X \leq 2)$?

$$\mathrm{P}(X \leq 2) = F_X(2) = f_X(-1) + f_X(1) + f_X(2) = 0.25 + 0.15 + 0.2 = 0.6$$

DHBW Duale Hochschule Baden-Württemberg Mannheim

Continuous Random Variables

- For now on, we consider a continuous random variable X on $\mathbb{R}$.
- Since there are infinite many values between any two points, it is poor to assign a probability value to a single point
- We rather introduce probabilities over intervals. How?

Continuous Random Variables probability density function

- For a continuous rv X with CDF $F_X(x)$, its derivative

$$f_X(x) \equiv \frac{\partial F_X(x)}{\partial x}$$

is called the probability density function (pdf).

- From above's equation we conclude

$$F_X(x) \equiv \int_{-\infty}^{x} f_X(u)\ \mathrm{d}u\ .$$

$\rightsquigarrow$ The pdf is a function for which the area under the curve corresponding to any interval is equal to the probability that X will take on a value in that interval

Continuous Random Variables pdf Properties

- $f_X(x) \geq 0 \ \forall x$ (since 'probability under curve')

  (*necessary* condition)

- $\int_{-\infty}^{\infty} f_X(x) \, \mathrm{d}x = 1$ (since $\lim_{x \to \infty} F_X(x) = 1$)

  (*necessary* condition)

- $\mathrm{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) \, \mathrm{d}x$

- $\mathrm{P}(X = x) = \int_x^x f_X(u) \, \mathrm{d}u = 0$

- $\mathrm{P}(a < X \leq b) = \mathrm{P}(a \leq X \leq b) = \mathrm{P}(a \leq X < b)$
  $= \mathrm{P}(a < X < b)$

Example: Uniform Distribution (Appetiser)

- A random variable X is uniformly distributed over $[a, b]$ (denoted by $X \sim U(a, b)$) if

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{else} \end{cases}.$$

- We consider $X \sim U(0, 2)$ with

$$f_X(x) = \begin{cases} \frac{1}{2} & x \in [0, 2] \\ 0 & \text{else} \end{cases}.$$

- What is $\mathrm{P}(0 \leq X \leq 1)$?

Example: Uniform Distribution (Appetiser)

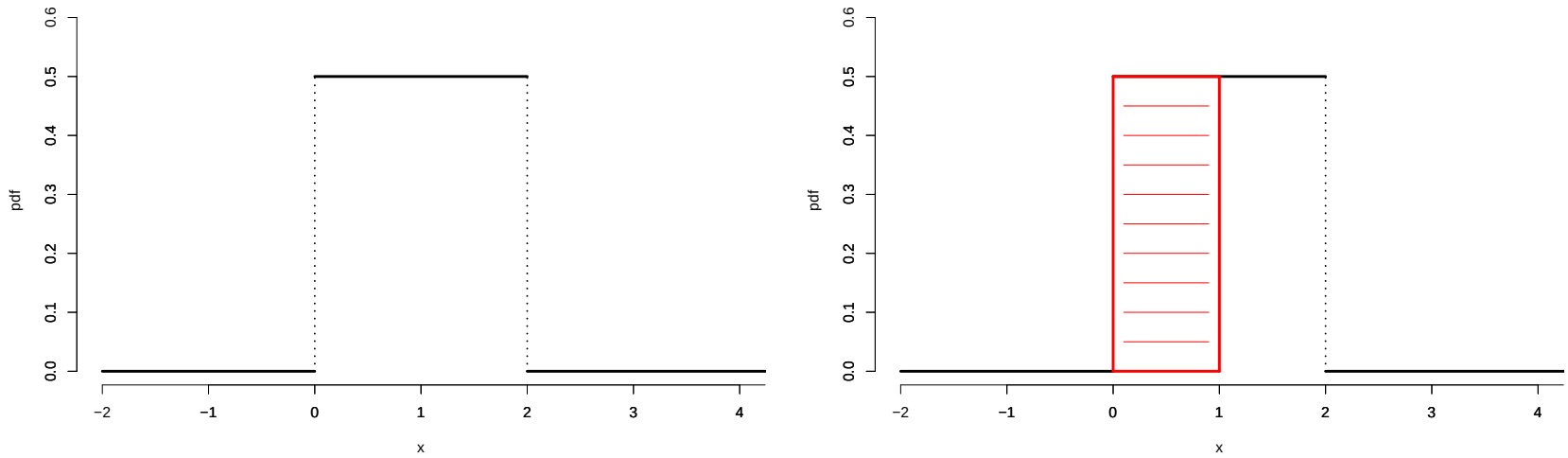

Figure: $f_X(x)$ for $x \sim U(0,2)$

$$\rightsquigarrow \mathrm{P}(0 \leq X \leq 1) = \int_0^1 f_X(x)\ \mathrm{d}x = \tfrac{1}{2} - 0 = \tfrac{1}{2}$$

Example: Check for Validity (1)

$$\text{Is } f_X(x) = \begin{cases} \frac{1}{2}x & 0 \leq x \leq 2 \\ 0 & \text{else} \end{cases} \quad \text{a valid pdf?}$$

- $f(x) \geq 0 \; \forall x$ ✓
- $\int_{-\infty}^{\infty} f_X(x) \, \mathrm{d}x = \int_0^2 \frac{1}{2}x \, \mathrm{d}x = \left[\frac{1}{4}x^2\right]_0^2 = \frac{1}{4}2^2 - \frac{1}{4}0^2 = 1$ ✓

$$\rightsquigarrow f_X(x) \text{ is a valid pdf}$$

Example: Check for Validity (2)

$$\text{Is } f_X(x) = \begin{cases} \sqrt{x} & 0 \leq x \leq 1 \\ 0 & \text{else} \end{cases} \quad \text{a valid pdf?}$$

- $f(x) \geq 0 \; \forall x$ ✓
- $\int_{-\infty}^{\infty} f_X(x) \, \mathrm{d}x = \int_0^1 \sqrt{x} \, \mathrm{d}x = \left[\frac{3}{2} \sqrt[3]{x^2} \right]_0^1 = \frac{3}{2} - 0 = \frac{3}{2} \neq 1$ ✗

$$\rightsquigarrow f_X(x) \text{ is not a valid pdf}$$

Example: Derive the CDF

Consider again the pdf $f_X(x) = \begin{cases} \frac{1}{2}x & 0 \leq x \leq 2 \\ 0 & \text{else} \end{cases}$. What is the corresponding CDF?

Since $\int \frac{1}{2}x \, \mathrm{d}x = \frac{1}{4}x^2$, the corresponding CDF is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4}x^2 & 0 \leq x \leq 2 \\ 1 & \text{else} \end{cases} .$$

Table of Contents

Table of Contents

Expected Value

Motivation: Looking for simple values to describe X, like a long-time average.

- Let X be a discrete rv with realisations x_i. The expected value (expectation/mean) is defined by

$$\mathrm{E}(X) \equiv \sum_i x_i \cdot f_X(x_i) \equiv \sum_i x_i \cdot \mathrm{P}(X = x_i) \ .$$

- Let X be a continuous rv. The expected value is defined by

$$\mathrm{E}(X) \equiv \int_{-\infty}^{\infty} x \cdot f_X(x) \ \mathrm{d}x \ .$$

- Notation: Commonly, the expected value of X is denoted by μ_X in statistics

DHBW Duale Hochschule Baden-Württemberg Mannheim

Expected Value (Advanced Definition)

Let $g(X)$ be a real valued function of a rv X.

- Let X be a discrete rv with realisations x_i. Then

$$\mathrm{E}(g(X)) \equiv \sum_i g(x_i) \cdot f_X(x_i) \equiv \sum_i g(x_i) \cdot \mathrm{P}(X = x_i) \ .$$

- Let X be a continuous rv. Then

$$\mathrm{E}(g(X)) \equiv \int_{-\infty}^{\infty} g(x) \cdot f_X(x) \ \mathrm{d}x \ .$$

Expected Value Example

- Let X be rolling a die. What is the expected value of X?

$$\mathrm{E}X = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5 \, .$$

- Let Y be the rv, where you square the result of rolling a die - since math is fun! What is $\mathrm{E}Y$?

  $Y = g(X) = X^2$ with $g(x) = x^2$. Therefore

$$\mathrm{E}Y = \mathrm{E}g(X) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = \frac{91}{6} \, .$$

- Let X be the gain when betting 1\$ on a single number playing French Roulette. What is $\mathrm{E}X$?

$$\mathrm{E}X = 35\$ \cdot \frac{1}{37} - 1\$ \cdot \frac{36}{37} = -\frac{1}{37}\$ \approx -0.027\$ \, .$$

DHBW Duale Hochschule Baden-Württemberg Mannheim

Expected Value Properties

For any constant c and any function $g(X)$ and $h(X)$ (for which the expected value exist), it holds that

- $\mathrm{E}(c) = c$
- $\mathrm{E}(c \cdot g(X)) = c \cdot \mathrm{E}(g(x))$
- $\mathrm{E}(g(X) + h(X)) = \mathrm{E}(g(X)) + \mathrm{E}(h(X))$
- if $g(X) \leq h(X) \Rightarrow \mathrm{E}(g(X)) \leq \mathrm{E}(h(X))$

Proof $\mathrm{E}(g(X) + h(X)) = \mathrm{E}(g(X)) + \mathrm{E}(h(X))$

From the definition of $\mathrm{E}(g(X))$, we see that

$$\begin{aligned}
\mathrm{E}(g(X) + h(X)) &\equiv \int_{-\infty}^{\infty} \big(g(x) + h(x)\big) f_X(x)\ \mathrm{d}x \\
&= \int_{-\infty}^{\infty} g(x) f_X(x) + h(x) f_X(x)\ \mathrm{d}x \\
&= \int_{-\infty}^{\infty} g(x) f_X(x)\ \mathrm{d}x + \int_{-\infty}^{\infty} h(x) f_X(x)\ \mathrm{d}x \\
&\equiv \mathrm{E}(g(X)) + \mathrm{E}(h(X))
\end{aligned}$$

Variance

Motivation: Looking for simple values to describe X, like the dispersion of a variable, its variance

$$\mathrm{Var}(X) = \mathrm{E}\big[(X - \mathrm{E}(X))^2\big] = \mathrm{E}(X^2) - \mathrm{E}(X)^2 \ .$$

- Let X be a discrete rv with realisations x_i and expected value μ_X. The variance is defined by

$$\mathrm{Var}(X) \equiv \sum_i (x_i - \mu_X)^2 \cdot f_X(x_i) \equiv \sum_i (x_i - \mu_X)^2 \cdot \mathrm{P}(X = x_i) \ .$$

- Let X be a continuous rv. The variance is defined by

$$\mathrm{Var}(X) \equiv \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f_X(x) \ \mathrm{d}x \ .$$

Variance

- Notation: Commonly, the variance of X is denoted by σ_X^2 in statistics

- Recap the definition of $\mathrm{E}(g(X))$ and set $g(X) = (X - \mathrm{E}(X))^2$ to get the expressions of

$$\mathrm{E}(g(X)) = \mathrm{E}\big[(X - \mathrm{E}(X))^2\big]$$

for the discrete and for the continuous case, respectively.

Variance Example

- Let X be rolling a die. What is the variance of X?

$$\mathrm{Var}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{546}{36} - \frac{441}{36} = \frac{105}{36} \approx 2.92 \,.$$

- Let X be the gain when betting 1\$ on a single number playing French Roulette. What is $\mathrm{Var}X$?

1. $\mathrm{E}X^2 = 35^2\$ \cdot \frac{1}{37} + (-1)^2\$ \cdot \frac{35}{37} = \frac{1261}{37}$.
2. $(\mathrm{E}X)^2 = \left(-\frac{1}{37}\right)^2$
3. $\mathrm{Var}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{1261}{37} - \left(-\frac{1}{37}\right)^2 \approx 34.08$.

DHBW Duale Hochschule Baden-Württemberg Mannheim

Variance Properties

For any constant c and any rv X, it holds that

- $\text{Var}(X) \geq 0$
- $\text{Var}(c) = 0$
- $\text{Var}(X + c) = \text{Var}(X)$
- $\text{Var}(c \cdot X) = c^2 \text{Var}(X)$

Standard Deviation

Motivation: Looking for simple values to describe X, like the dispersion of a variable in the same unit scale, its standard deviation

$$\sigma_X \equiv \sqrt{\sigma_X^2} \equiv \sqrt{\text{Var}(X)}$$

- The standard deviation is a scaled version of the variance, more appropriate to give inferences about X

Table of Contents

Table of Contents

Binomial Distribution (Binomial Coefficient)

Motivation: Consider Scrooge McDuck chooses two of his grandnephews (Huey, Dewey and Louie) to count his money.

How many combinations of his grandnephews are possible? (disregarding the order)

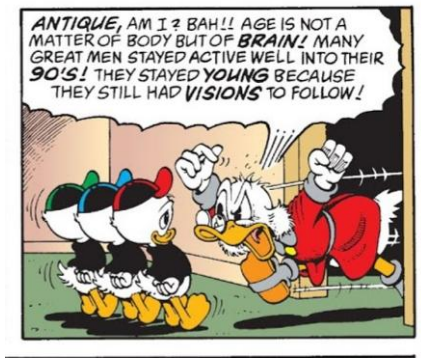

Figure: Source: https://i.redd.it/156jmpj0w9m21.jpg

1. Huey & Dewey
2. Huey & Louie
3. Louie& Dewey

Binomial Distribution (Binomial Coefficient)

Motivation: Consider you want to choose k distinct objects among n objects, disregarding the order. How many combinations are possible?

- We have n choices for the first pick
- We have $n-1$ choices for the second pick
- We have $n-2$ choices for the third pick
- We have ...

$\rightsquigarrow$ We have $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)$ many possible arrangements

Binomial Distribution (Binomial Coefficient)

Scrooge McDuck had 6-many possible arrangements:

1. Huey & Dewey
2. Dewey & Huey
3. Huey & Louie
4. Louie & Huey
5. Louie& Dewey
6. Dewey & Louie

But since he doesn't mind about the arrangements, he has to divide by the amount of ways in which his grandnephews can be arranged: By 2

$\rightsquigarrow 6/2 = 3.$

Binomial Distribution (Binomial Coefficient)

Dividing by the amount of ways the picks can be arranged:

- 1 pick $\rightsquigarrow$ divide by 1
- 2 picks $\rightsquigarrow$ divide by $2 \cdot 1$
- 3 picks $\rightsquigarrow$ divide by $3 \cdot 2 \cdot 1$
- 4 picks $\rightsquigarrow$ ...

- k picks $\rightsquigarrow$ divide by $k \cdot (k-1) \cdot \ldots \cdot 1 \equiv k!$

We get

$$\frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \equiv \binom{n}{k}$$

which is called the binomial coefficient. The binomial coefficient gives us the amount of combinations possible when choosing k-many distinct objects among n-many different objects.

Binomial Coefficient (Example)

Playing Poker, how many five-card hands are possible?

$$\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot \cdot 1} = 2\ 598\ 960\ .$$

Binomial Distribution

Setup for a rv X: What is the probability of getting exactly k-many elements of interest in n experiments with replacement, where p is the probability of having success and $q = 1 - p$ is the probability of failure.

This probability is given by the Binomial Distribution with pmf

$$f_X(k) = \binom{n}{k} p^k q^{n-k}, \; k = 0, 1, 2, \ldots, n \; .$$

Recap: $f_X(k) = P(X = k)$

Binomial Distribution Interpretation

- $p^k q^{n-k}$ probability of having success
- $\binom{n}{k}$ amount of ways to have success

Binomial Distribution Example

Consider there is a lake with 100 salmons and 200 sardines. You want to fish seven fish today, while you throw every fish back into the lake. What is the probability of fishing exactly four salmons (when fishing seven fish overall)?

- $p = \frac{1}{3}$ (probability of having 'success')
- $q = 1 - p = \frac{2}{3}$ (probability of 'failure')
- $k = 4$
- $n = 7$

$$f_X(4) = \binom{7}{4} \left(\frac{1}{3}\right)^4 \left(\frac{1}{3}\right)^3 \approx 0.1280 = 12.8\%$$

DHBW Duale Hochschule Baden-Württemberg Mannheim

Binomial Distribution: Characteristics

Let X be a Binomial distributed rv with n independent experiments (with replacement) and success probability p. We write $X \sim B(n, p)$ and it holds

- pmf:

$$f_X(k) = \binom{n}{k} p^k q^{n-k}, \ k = 0, 1, 2, \ldots, n \ .$$

- CDF:

$$F_X(k) = \mathrm{P}(X \leq k) =\equiv \sum_{i}^{i:i \leq k} f_X(i) = \sum_{i}^{i:i \leq k} \binom{n}{i} p^i q^{n-i} \ k = 0, 1, 2, \ldots, n \ .$$

- expected value:

$$\mathrm{E}(X) = np$$

- variance:

$$\mathrm{Var}(X) = np(1-p)$$

Binomial Distribution Example

Consider there are 60 students participating in this course. Each student has probability 0.8 to pass the exam.

Q1: What is the probability that at least 50 students pass the exam?

Q2: How many students pass on average?

It is reasonable that the students perform independently. We then have

- $p = 0.8$ (probability of having 'success')
- $q = 1 - p = 0.2$ (probability of 'failure')
- $n = 60$

Q1
$$\begin{aligned} & \mathrm{P}(X \geq 50) = 1 - \mathrm{P}(X < 50) = 1 - \mathrm{P}(X \leq 49) = F_X(49) \\ = & \sum_{i=0}^{49} \binom{60}{i} 0.8^i \cdot 0.2^{(60-i)} \approx 1 - 0.6766 = 32.34\% \end{aligned}$$

Q2
$$E(X) = n \cdot p = 60 \cdot 0.8 = 48.$$

DHBW Duale Hochschule Baden-Württemberg Mannheim

Geometric Distribution

Setup for a rv X: What is the probability that the k-th experiment is the first of having success, with replacement, where p is the probability of having success and $q = 1 - p$ is the probability of failure.

This probability is given by the Geometric Distribution with pmf

$$f_X(k) = q^{k-1}p, \quad k = 1, 2, \dots \quad .$$

Geometric Distribution Example

Consider you are on a party and you are going to the kitchen. In front of you, there are walking four persons towards the kitchen. There is a probability of 0.3 that each person goes to the fridge and a probability of 0.7 of not going to the fridge (and playing Beer-Pong rather).
What is the probability, that only the person in front of you is going to the fridge?

- $p = 0.3$ (probability of having 'success')
- $q = 0.7$ (probability of 'failure')
- $k = 4$

$$f_X(4) = 0.7^3 \cdot 0.3 \approx 0.1029 = 10.29\% \ .$$

DHBW Duale Hochschule Baden-Württemberg Mannheim

Geometric Distribution: Characteristics

Let X be a Geometric distributed rv under independent experiments (with replacement) and success probability p. It holds

- pmf:

$$f_X(k) = q^{k-1}p, \quad k = 1, 2, \ldots$$

- CDF:

$$F_X(k) \equiv \sum_{i}^{i:i \leq k} f_X(i) = 1 - q^k, \; k = 1, 2, \ldots$$

- expected value:

$$\mathrm{E}(X) = \frac{1}{p}$$

- variance:

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

Geometric Distribution Example

Fighting with Haunter, you want to use Hypnosis to make the foe falling asleep. Hypnosis has an accuracy of 60%.

Q1: What is the probability that Haunter needs more than two attacks to accomplish Hypnosis?

Q2: On average, how many tries does Haunter need?

Figure: Source: vizzed.com

- $p = 0.6$ (probability of having 'success')
- $q = 0.4$ (probability of 'failure')

Q1 $\mathrm{P}(X > 2) = 1 - \mathrm{P}(X \leq 2) = 1 - F_X(2) = 1 - (1 - 0.4^2) = 16\%$.

Q2 $\mathrm{E}(X) = \frac{1}{0.6} = \frac{5}{3} \approx 1.66$.

Geometric Distribution Example

Imagine you can dope your Pokémon to increase its accuracy for conducting Hypnosis.

Q1: Until which accuracy level do you have to dope Haunter in order to achieve an accomplish-probability of 95% after three attacks?

Q2: On average, how many tries does your doped Haunter need now?

Q1 $\quad \mathrm{P}(X \leq 3) = F_X(3) = 1 - q^3$.

$\rightsquigarrow \quad 1 - q^3 \overset{!}{=} 0.95 \quad \Rightarrow \quad q \approx 0.3684 = 36.84\%$.

$\rightsquigarrow \quad p = 1 - q = 1 - 0.3684 = 0.6316 = 63.16\%$.

Q2 $\quad \mathrm{E}(X) = \frac{1}{0.6316} \approx 1.58$.